DOCUMENT RESUME

ED 132 055 SE 021 755

AUTHOR Wollman, Warren
TITLE Intellectual Development Beyond Elementary School VI, Controlling Variables: A Survey.
INSTITUTION California Univ., Berkeley. Lawrence Hall of Science.
PUB DATE [76]
NOTE 15p.; Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Cognitive Ability; *Cognitive Development; Conservation (Concept); *Educational Research; *Elementary Secondary Education; Intellectual Development; Mental Development; *Science Education
IDENTIFIERS *Piaget (Jean); Research Reports

ABSTRACT

Reported is a study of 1555 students, grades 4-12, regarding understanding the concept of controlling variables. Subjects were examined on a task involving spheres rolling down an incline and striking a target sphere; they were evaluated upon their explanations of answers, not their choice of answers. Findings indicated that the Piagetian stage of formal thinking does not develop spontaneously; however, for above average subjects, the beginning of a formal stage usually develops by age 14, and somewhat later for less able subjects. Also, it is probable that there is no single formal concept of controlling variables. A copy of the evaluation instrument is included. (SL)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

6-10-76

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED132055

Intellectual Development Beyond Elementary School VI
Controlling Variables: A Survey

Warren Wollman
AESOP*, Lawrence Hall of Science
University of California, Berkeley

The concept of a controlled experiment is essential to what Dewey has called critical or reflective thinking (Dewey, 1971). Piaget (Inhelder and Piaget, 1958) believes that the development of the ability to control variables is an index of intellectual development in general. According to Piaget, this ability should normally be developed by around age 15. However, various researchers, using Piaget's criteria, have observed that a very substantial fraction of adults have not acquired this ability (see, e.g., Renner and Lawson, 1973; Kohlberg and Gilligan, 1971). Whether or not Piaget is correct in asserting that the concept of controlling variables is spontaneously acquired in the normal course of affairs, it is desirable that it be acquired. Indeed, it serves critical thinking not only in the sciences, but also in the evaluation of evidence in history, current events, and economics where direct experimentation is not possible.

What can educators do to promote acquisition of the concept? Piaget believes that only certain experiences can be effective and that even these are limited. He makes a distinction between learning and development. Briefly, he asserts that expository teaching cannot substantially accelerate the acquisition of certain fundamental concepts, e.g., that of controlling variables, because the development of these concepts depends on maturational factors. Yet, development requires interaction with the environment. Thus the teacher should provide opportunities for such interactions. But of what sort?

Commenting on the potential effectiveness of a classroom demonstration of the technique of controlling variables, Piaget said, "It would be completely useless. The child must discover the method for himself." (quoted by Hall, 1970). In fact, even when a method is discovered in some contexts, e.g., comparing the bounce of two tennis balls under a variety of novel but equal conditions, the child may not be able to state nor even be aware of his general method per se, namely assuring equal testing conditions. When faced with new problems, the child will begin anew, not perceiving the applicability of his prior successful solutions to other problems requiring controlled experiments (Wollman, 1975; Lawson and Wollman, 1975). The child constructs particular solutions and represents them in particular, i.e., context-dependent terms. For example, after rejecting a variety of unfair comparisons of tennis balls proposed by the experimenter, the child "summarizes" the characteristics of fair comparisons by stating that "you must drop them from the same height and let them go the same way--don't throw one and not the other--and let them bounce on the same kind of floor" (the experimenter had used unequal heights, modes of release, and floor surfaces). The child does not simply say something equivalent to, "Compare them in the same way." Even when repeatedly asked to "say it in one short sentence without mentioning anything like height and the floor," or when asked to "say what it is about all unfair (or fair) comparisons that make them unfair (or fair)," otherwise articulate students (age 12-13) of average intelligence are usually at a loss for words (Lawson and Wollman, 1975).

---

*AESOP (Advancement of Education in Science-Oriented Programs) is supported by a grant from the National Science Foundation.

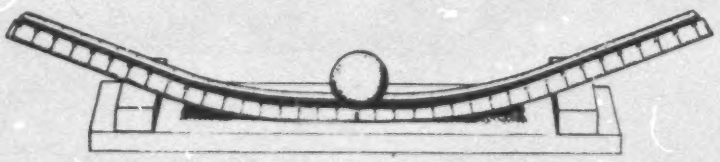
SE 021 755

Thus, different levels of ability are revealed by adequate solutions to tasks requiring controlled variables. Piaget's tasks are probably too demanding to reveal the lower level abilities which develop prior to the mature concept but which do suffice for the tennis balls task (though not for others in general). Rather than ascribe lack of general success to maturational factors and delay the introduction of the concept until the student is "ready" (as suggested by Elkind, 1972), it would be far more profitable to inquire as to the roots of the concept of controlling variables. A knowledge of these roots would aid educators in deciding what sorts of interactions should be provided a student in order that he "discover the method for himself." The tasks reported on in this survey were designed to be part of a determination of the developmental roots of the concept of controlling variables. The tennis balls task, though useful, was considered too "everyday" for the age range studied (9 years and older). Piaget's tasks, e.g., the bending rods and pendulum tasks (Inhelder and Piaget, 1958), were considered too demanding for a variety of reasons; they involved too many variables (Case, 1975) and phenomena that was too unfamiliar. (As a means of assessing the mature concept, Piaget's tasks suffice.) Tasks were designed around few and familiar variables in an understandable yet not "everyday" context.

The Tasks

Pilot studies confirmed that starting at least as early as the third grade, subjects (Ss) are familiar with the roles of the variables of weight and height in the following: a sphere rolls down an incline and strikes a target sphere, sending it up a second incline (see Figure 1). All tasks refer to the general role of a variable, e.g., the higher the starting position of the first sphere, the farther it will hit the target (other conditions being equal). Of course, the parenthetical remark is tacit. The tasks were meant to reveal to what extent S was aware of it.

FIG. 1



In one task, S must decide whether to release heavy spheres, light ones, or a combination in order to determine what happens when the sphere starting position is varied (the target weight is constant and given). In a second task, a simpler version of the first, S must decide whether to release a light or a heavy sphere from a high position, given that a light sphere was released from a low position. Again, the point is to see whether changing the position makes any difference in what happens to the target (other conditions being equal). The second task offers S two alternatives while in the first task, S must choose from four. Thus, the first task

should be more difficult than the second if S does not have a sufficiently powerful procedure for classifying alternatives. In fact, even when S is concerned with making an unambiguous comparison, if there are many variables to control then he may overlook some for lack of a sufficiently general method (Case, 1975). This point will be taken up again in the discussion section.

The first and second tasks require S to specify both terms, then only one term of a comparison respectively. In a third task, both terms are specificied in advance. S must simply judge whether the given (uncontrolled) experiment proves a given (general) assertion. In all three tasks, there are only two variables of concern (a minimal number), the one to be varied, e.g., sphere starting position, and the one to be held constant, e.g., sphere weight.

Subjects

To permit longitudinal comparisons, homogeneous samples were chosen. Ss were chosen from urban and suburban neighborhoods in the San Francisco Bay Area. While all the suburban Ss were drawn from the same community, this was not practical for the urban Ss. A wide variety of urban junior high schools feed into the urban high school chosen for the study and the urban 10th and 12th grades include Ss of much lower SES than of those Ss in the pre-high school urban grades. Table 1 gives the number of Ss from each grade.

Table 1. Number of Ss responding to questions 1, 2, or 3 in each grade for the urban and suburban schools.

Grade	4	6	7	8	9	10	12
Urban	54	119	185	272	228	62	52
Suburban	41	54	161	158	83	63	23

Procedure

Following a spoken introduction, classes responded to written questions. (It is understood that personal interviews might result in different performance.) Most Ss finished these questions in less than ten minutes. During the introduction, Ss observed a sample event and were asked to name the "things" affecting how far the target went. It was specifically pointed out to the class that the weights of the sphere and target as well as the starting position of the sphere made a difference in how far the target goes. Ss were asked to think of a way to prove (to someone who did not already know) that starting position "makes a difference in how far the target goes." After a pause and without asking for oral responses, three ways were suggested for evaluation: compare what happens starting from high and low positions using 1) two heavy spheres, one from each position, 2) two light spheres, 3) a combination of light and heavy. Then the first two questions were read aloud. At this point, the written test began.

The Written Test

1. Suppose you want to find out whether it makes a difference to release a sphere from the high or low position. Look at drawings (a) and (b) and decide which spheres you would release from the high and low positions in order to find out whether the position makes a difference.
   (i) Which sphere would you release from the high position? _____
   (ii) Which sphere would you release from the low position? _____.
   Please explain your answer. (Space was provided)

2. Suppose in drawing (b) someone releases a light sphere from the low position. Then which sphere would you release from the high position in order to see whether the position makes a difference.
   Light ____ Heavy ____. Please explain your answer. (Space provided)

3. Suppose there are two kinds of metals, A and B. You want to see which kind of metal is better for hitting a target far. You do the experiments in drawings (c) and (d).

   Suppose metal B hits its target farther than metal A hits its target. Does this prove metal B can hit a target farther than metal A can?
   Yes _____ No _____ Please explain your answer. (Space provided)

Scoring of Questions 1 and 2

The explanations were judged, not the choice of sphere weights. A scoring system was worked out which resulted in consistent agreement (87%) between two scorers and only minor disagreements as a rule. Four response levels were distinguished.

Level 1. All manifestly incorrect responses are included here. The most frequent one focussed entirely on the desirability of hitting the target far (see ex. 1.1, 1.2). Other responses confounded variables (ex. 1.3). Still others were incoherent (ex. 1.4). Some were (apparently) beside the point (ex. 1.5, 1.6).

Examples:

1.1	(i)	High-Heavy	"Heavy hits with more force."
	(ii)	Low-Heavy	
1.2	(i)	High-Light	"Light sphere from high position so it can build
	(ii)	Low-Heavy	up speed. Heavy can hit far from any position."
(to question 2)			
1.3		Heavy	"The light one would not make it go very far while the heavy one would make it go very high."
1.4	(i)	High-Heavy	"Because when you let go of a heavy sphere you
	(ii)	Low-Heavy	want it to go high on the ramp so a heavy sphere will push a light sphere and make it go high because it's light."
1.5	(i)	High-Light	"A light sphere would need space to move the target
	(ii)	Low-Heavy	so you would place it at the high position. A heavy sphere would not need as much speed because of its weight so you could place it at the low position."
1.6	(i)	High-Heavy	"A light sphere might not even move the target.
	(ii)	Low-Heavy	You could tell better with a heavy sphere."

Level 2. This level includes two principal kinds of response. One kind "explains" the choice of equal weights either by describing a procedure for using them or by predicting the outcome of the experiment, usually the latter (ex. 2.1, 2.2). The other "explains" the choice of a light sphere from high vs. a heavy sphere from low by stating that the light sphere "needs to be high" to make up for its lack of weight or by predicting (nearly) equal distances reached by the target both times (ex. 2.3, 2.4). Note that since there is no information available to S which could indicate the extent of compensation one might expect, these subjects fail to anticipate a potential ambiguity. All responses of this type will be referred to as compensation (C) responses. Lower level and implicit C's were assigned to Level 1 (see ex. 1.2, 1.5). Also included in Level 2 are ambiguous or incomplete responses which received the benefit of a doubt.

Examples:

Examples:

2.1	(i)	High-Heavy	"Heavy high because want to see how far it will
	(ii)	Low-Heavy	go. Heavy low because then you'll find out how much higher it goes from the high position."
2.2	(i)	High-Heavy	"To see if the starting point makes a difference,
	(ii)	Low-Heavy	let a heavy ball go from a high point and see how far it goes up the other side and then a heavy one from a low position and compare your results."
2.3	(i)	High-Light	"At the high position the sphere would get more
	(ii)	Low-Heavy	speed and would be able to hit the target about as far as the heavy sphere at the low position where the ball wouldn't get as much speed but since it is bigger it would have more power."

2.4	(i)	High-Light	"You could place the light sphere high because
	(ii)	Low-Heavy	the extra height could make up for its being lighter. This should even out the impact." (Compare 2.3 and 2.4 with 1.2 and 1.5)

Levels 3 and 4. Using equal weights, these explanations indicate an overt awareness of the idea of a comparison and the normative character of their procedure by expressing either the necessity or the desirability of equal weights or by implying that the equality of weights follows from the fact that different positions are used. Only rarely do these explanations also appeal to the physical outcome of the experiment. The focus has shifted from phenomena and the mechanics of experimental conditions to the logical aspect of procedure.

These responses often reject the ambiguity inherent in confounding weight and height. Some of these responses are clearly derived from a general realization that comparisons should not confound variables.

The distinction between levels 3 and 4 is primarily one of degree. Level 3 responses while expressing a preference for using the same weight and criticizing unequal weights, are less insistent than level 4 responses. Where a level 3 response uses "should," a level 4 response uses "must." The level 4 responses were judged as more definitive in the choice of equal weights. Typically, they first stated necessity, then stated that unequal weights would preclude the possibility of a comparison. Also included in level 4 were terse responses referring to "experimental controls."

Examples:

3.1	(i)	High-Heavy	"If you are using different positions, then you
	(ii)	Low-Heavy	would use the same weight."
3.2	(i)	High-Heavy	"I would use two of the same weight, with both
	(ii)	Low-Heavy	spheres heavy, the target would go further and you could see the difference better." (See ex. 1.6.)
3.3	(i)	High-Heavy	"I would use the same weight because different
	(ii)	Low-Heavy	weights cause different reactions."
3.4	(i)	High-Heavy	"Both of them should be the same. I don't think
	(ii)	Low-Light	the light would make it go far enough to really tell." (See ex. 1.6.)
(to question 2)			
3.5		Light	"If you use the same thing then you can tell better. A heavy ball wouldn't show you much."
(to question 2)			
3.6		Light	"A heavy sphere would roll harder than a light one."
4.1	(i)	High-Heavy	"In order to prove there is a difference you have
	(ii)	Low-Heavy	to use the same weight. If you use different weights there'd be no comparison." (See ex. 3.1.)
4.2	(i)	High-Light	"It's got to be the same weight because if they
	(ii)	Low-Light	were different, the rolling of the target would be different even if the spheres were released from the same position." (See ex. 3.3.)

4.3	(i)	High-either	"You must use the same thing, otherwise the experiment is useless." (See ex. 3.4.)
	(ii)	Low-either	
4.4	(i)	High-Light	"Spheres should be the same in both since height is different. If too many things are different, you can't tell which things make a difference."
	(ii)	Low-Light	
(to question 2) 4.5		Light	"If you used the heavy one that wouldn't prove anything." (See ex. 3.5)
(to question 2) 4.6		Light	"Of course heavy would make a difference. So you must use light." (See ex. 3.6.)

Scoring of Question 3

Responses were classified simply as either acceptable or unacceptable. Most of the acceptable responses were quite brief. Many Ss gave full explanations to the previous questions but were terse in rejecting the invalid proof in question 3, stating "the targets were different," or "a heavy target wouldn't go as far." Some went on to specify that "you have to have the same target to see which hit farther since naturally the light target would go farther." A very few pointed out that the metal spheres "might not be the same weight." Finally, a handful overlooked the unequal target weights and concluded that "the materials and the positions are the same so this proves it." These responses were accepted.

The unacceptable responses were usually attempts to explain the evidence in terms of an uncontrolled variable! Some examples are:

(a) (Yes) "Because target B is a light target and target A is heavy."
(b) (Yes) "Because it is hitting a lighter target that would make it go farther."
(c) (Yes) "Because metal B is heavier."
(d) (Yes) "Yes, if B was a heavier metal."

There were some responses which simply repeated the evidence with not a word of further explanation, much less of criticism.

(e) (Yes) "Because metal B hit it farther."
(f) (Yes) "Because you said so."

In addition, there were a number of responses which were either incoherent, made no reference to the evidence, or were without an explanation.

(g) (No) "Same as B."
(h) (No) "Because it could just be the surface or something."
(i) (No) "You have to try it more than once."

Results

The distribution of scores for questions 1 and 2 for all Ss is given in a contingency table (Table 2a). Question 1 was more difficult than 2. In Table 2b, the six principal entries of Table 2a are combined into four groups (the numbered cross-hatched areas). The numbers indicate four response "stages" based on both responses. The other numbers in Table 2b indicate the assignment of the non-principal entries to one of these stages (the meaning of

"stage" is clarified in the discussion). In other words, it is conjectured that the non-principal entries are "noise" while the grouped principal entries represent the "signal." The distribution of Ss in these stages is shown for each grade in Table 3.

Table 2 (a): Contingency table of responses to questions 1 and 2. Entries give % of total number of Ss (N = 1421).

		Question 2 Levels				
		4	3	2	1	Total on Question 1
Question 1 Levels	4	11%	.84	.07	-	11%
	3	2.8	7.0	.14	-	10
	2	2.4	9.9	11	2.5	26
	1	.63	4.2	8.4	39	53
Total on Question 2		16%	22	20	42	(1421)

(b) Cross-hatched areas show principal entries of Table 2A grouped into 4 response stages. Numbers indicate classification into stages for all entries.

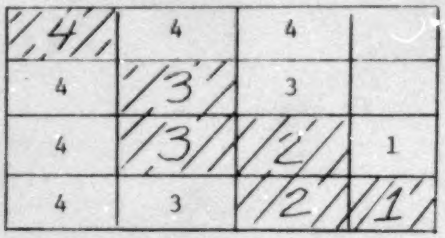


Table 3. Percent of (N) Ss in each grade at each stage. The 4th graders were classified on the basis of question 2 only.

	Urban Grades						Suburban Grades					
	4	6	7	8	9	10&12	4	6	7	8	9	10&12
Stage 4	-	-	4.9	14	18	12	5.1	14	20	25	32	46
3	-	9.2	13	20	24	16	15	27	20	30	36	24
2	1.9	15	14	20	16	31	23	8.1	25	22	16	21
1	98	76	69	45	42	41	56	51	34	22	16	9.4
(N)	54	98	185	272	228	114	39	37	161	158	83	85

Overall, there is a gradual trend with increasing grade from (1) an early focus on simply "hitting the target far" in some global or absolute sense, to (2) relating "far" to the position varible, with the outcome of the experiment justifying the choice of weights, to (3) a focus on antecedent conditions as jusifying this choice, to (4) a (relatively) definitive exclusion of logical ambiguity. Stages 3 and 4 indicate at least some explicit awareness of the idea of a comparison in the sense of a controlled experiment.

Virtually all the urban 4th graders were in stage 1 (based on their response to question 2--question 1 proved too difficult for them to even attempt it). By the 9th grade, 42% (94 of 228) were in stage 3 or 4. The urban high school shows substantially the same distribution for the 10th and 12th grades with only 28% (32 of 114) in stage 3 or 4 (recall that the high school population differs from the other urban samples).

In the suburban community, the percentage in stage 3 or 4 goes from 21% (8 of 39) in the 4th grade (based only on a response to question 2) to 68% (57 of 83) in the 9th grade. For the combined 10th and 12th grades, the figure is 70% (59 of 85) with no appreciable difference between the two grades. The relative contribution to these figures of stage 4 increases at higher grades (especially at the 10th and 12th, which may have been due to course work in an experimental science course; most of the responses explicitly mentioning "experimental controls" came from these two grades). Note that almost half of the 4th grade Ss showed some understanding of question 2.

The urban grades 7-9 were tested in tracked math classes. The distribution in stages of the combined grades 7-9 for the different tracks is shown in Table 4. The four tracks are easily distinguished. Most of the gains from the 7th to 9th grades were apparently made by Ss in tracks 3 and 4. A longitudinal comparison is only approximate, however, because the criteria for distinguishing the tracks differ in each grade and Ss

move in and out of tracks. A similar observation was made of the development of the ratio concept (Karplus and Karplus, 1972). It may be that for a large part of this population, there is little development of these concepts beyond the elementary school.

Table 4. Percent of (N) Ss for each math track at each stage (for the combined urban grades 7, 8, and 9).

Math Track	4	3	2	1
Stage 4	38%	11	4.0	-
3	35	27	11	-
2	17	23	16	9.0
1	10	39	69	91
(N)	(143)	(200)	(275)	(67)

Either with increasing grade or with higher math track (within a given grade), there is (1) an increase in the proportion of stages 3 and 4, accompanied by (2) a decrease in the proportion of stage 1.

Of special interest are the numerous compensation (C) responses to question 1 (see ex. 2.3 and 2.4, also 1.2 and 1.5). The low level C's (ex. 1.2) were usually concerned only with hitting the target for (like ex. 1.1). Many others (ex. 2.3) predict the target will go about as far (or be hit about as hard) whether hit by light from high or heavy from low. If this occurred and if the role of weight is presumed known, then a valid inference could be drawn concerning the position variable. Only a few argued explicitly in this manner. The intent of the remainder was unclear.

The C's did provide some positive information by their absence! About thirty Ss spontaneously crossed out a C response and then chose equal weights, usually stating that with equal weights it is easier to tell whether position makes a difference. None switched from equal weights to C. Also, the frequency of the C's falls off in faster math tracks and in higher grades. These results confirm the wholesale exclusion of C's from levels 3 and 4.

Summing to this point, the results allow one to conjecture a developmental trend:

Stage 1: No display of the concept of a comparison; focus on 'far' in an absolute sense.

Stage 2: An intuitive, inarticulate and concrete use of the concept with focus on the mechanics and/or the result of the comparison.

Stage 3: A transitional stage (including many level 2 C's) with recognition of the desirability of equal weights but without the "logical imperative" of doing so.

Stage 4: Definite rejection of unequal weights as a source of logical ambiguity with focus on the logical consequences of antecedent conditions.

The 3rd question was less challenging than the other two, as expected (see Table 5). At the 4th grade, 44% (18 of 41) of the suburban Ss, but only 5.6% (3 of 54) of the urban Ss rejected the uncontrolled experiment in question 3. By the 6th grade, the percentages rose to 54% and 30% respectively. Most the remainder (at all grades) explained why one sphere went further than the other, rather than with an evaluation of the procedure itself. By the 10th grade, the question was trivial for the suburban Ss while about two-thirds of the urban sample responded acceptably. In neither the urban nor suburban high school did the 12th grade score higher.

Table 5. Percent of (N) Ss responding acceptably to question 3 for each grade.

Grade	4	6	7	8	10&12
Urban	5.6 (54)	30 (90)	-	64*(70)	65 (88)
Suburban	44 (41)	54 (48)	65 (63)	64 (58)	91 (68)

*The entry for the urban 8th grade was estimated using a weighted average of a "gifted," a "fast" and an "average" math class.

Responses to question 3 confirm the division into stages of Table 2b. Table 6 displays for each stage and grade the percentage of acceptable responses to question 3. Stages 1 and 2, and a combined 3 and 4 are differentiated by question 3. Evidence for the distinction between stages 3 and 4 is provided by the fact that of those Ss who give the C response to question 1, far more respond to question 2 at level 3 than at level 4. If there were no difference between these levels, then C's would respond to question 2 at level 3 as often as at level 4.

Table 6. Percent of (N) Ss responding acceptably to question 3 for each stage in each grade (except the 4th). Final column gives % of (N) Ss for each stage summed over urban and suburban grades.

	Urban			Suburban				All
	6	8	10 & 12	6	7	8	10 & 12	
Stage 4	-	100 (16)	100 (14)	100 (4)	94 (16)	91 (11)	100 (35)	98% of (96)
3	100 (6)	100 (22)	86 (14)	63 (8)	85 (13)	79 (14)	91 (22)	87% of (99)
2	63 (8)	64 (14)	85 (27)	60 (5)	67 (12)	62 (13)	63 (8)	70% of (87)
1	29 (55)	47 (17)	25 (32)	33 (18)	33 (24)	33 (18)	50 (2)	32% of (166)

Unfortunately, not everyone was asked question 3 because it was expected to be too easy. It was not and so it did provide additional information about the division into stages. Also of those stage 1 Ss who answered it, about one-third responded acceptably. On the basis of all three questions, one expects more than 90% of the suburban 10th grade Ss to show at least some understanding of a controlled experiment while the urban figure is about 75%. Using Piaget's tasks and criteria, the overall figure is only about 50% (Kohlberg and Gilligan, 1971).

Discussion

It is likely that the concept of a controlled experiment cast in the idea of fair comparison begins to develop early, perhaps near the start of what Piaget calls the concrete stage of development (6 years) (Case, 1975; Wollman 1975; Piaget, 1932). Fairness and evenness are surely at the root of the mature concept. But it is also likely that Ss remain unaware of general criteria of fairness even when they are capable of correctly judging a variety of comparisons as fair or unfair (Lawson and Wollman, 1975). Also lacking at the concrete stage is a clear idea of what to do in general when asked to judge a (possibly) multiple-caused event. Ss do not attempt to first analyze the event in terms of a complete set of variables. Even when these are specified, they do not then systematically determine the roles of the variables by varying them one at a time. In circumstances where experimentation is not possible, as with historical or current events, they do not go beyond the event to ask for other information that might clarify the (hypothesized) roles of variables (Peel, 1971).

Piaget's formal stage is supposed to remedy this situation (by around age 15). However, there is no reason to believe it develops spontaneously in a majority of Ss. The results of this study indicate that for above average Ss, a fairly clear idea of a fair comparison (but not necessarily a mature "formal stage" idea) develops by age 14 on the average. For less able Ss, there is rather less development of the concept. In fact, most of the gains from the 7th to 9th grades at the urban school were made by the relatively few Ss in the faster math tracks 3 and 4.

The three questions indicate three levels of difficulty. Question 3 required only that S note the unevenness of a comparison. As noted before, this is possible even when S cannot specify in general what constitutes or is preferable about an even comparison. Tasks such as question 3 in which all the information is provided might be a good beginning towards developing a critical attitude, particularly for younger Ss. Making unfair comparisons fair could follow, then completing comparisons as in question 2. The real task for S would be to invent a single way (an algorithm) to describe how to make fair comparisons, after having solved a variety of tasks. Without a general procedure for assuring that a comparison is fair and leads to unambiguous results, S would have to start almost anew with each new problem. In particular, if many unfamiliar variables were involved, S would not be able to keep track of them without a general, hence powerful procedure.

Question 2 required S to choose from two alternatives, question 1 from four. Increasing the number of alternatives would make no difference to one who realized that the only concern is with eliminating weight differences. Without this anticipatory guiding idea to act as a censor, more choices mean more to think about and thus S is led to unnecessarily consider choices

which might prove to be inconclusive. The compensation response to question 1 was one such choice. At best, it is a reasonable one. However, it is more complicated than the simple expedient of a controlled experiment (as some Ss explicitly noted).

In fact, the logical clarity of controlling variables is hypothesized by Piaget to underly its psychological development (in analogous fashion to its historical development). According to Piaget, S must first experience logical confusion before a way out is to develop. Moreover, a variety of such experiences is necessary if S is to develop a general procedure for arriving at unambiguous judgments by separating variables with controlled experiments. Generality results from giving a unified representation of the various experiences.

What if tasks of this kind were to become as "everyday" as comparing how two tennis balls bounce? Or as common as simple recall tasks are in our schools? In fact, such tasks are quite rare. Yet interviews (rather than written tests) with above average 3rd graders revealed an inarticulate preference for controlling variables on the spheres question 1. There is every reason to believe that the intuitive, concrete notions of fairness and evenness, restricted at first to particular situations involving few and familiar variables, would soon develop into a clear and general working concept if only Ss were confronted by the question of validity of judgments. Whether this development is achieved by 10 years or by 15 years is relatively unimportant. It is more important to realize that (1) development can be nurtured early in elementary school, (2) many contexts will do although science offers the simplest situations (see, e.g., the SCIS units Interaction and Systems and Subsystems and Variables), (3) without a wide variety of experiences, the general concept of a controlled experiment is not likely to develop.

Finally, as for the Piagetian notion of "stage," it is very unclear whether this notion was meant to apply to individuals. Like Goethe's "Idea" of a plant, Piaget's theoretical stage concept is meant to encompass all individuals. Hence, it ignores even substantial individual differences as well as inconsistencies in the behavior of one individual. At best, it is a concept of competence, not performance. Lunzer (1973) has pointed out how difficult it is to define a formal stage, theoretically or empirically. In the present study, the term "stage" was only used to describe moments in a development that is gradual and continuous. Instead of the dichotomy concrete/formal, it would be preferable to think of a continuum from relatively concrete to relatively formal. One tentative empirical definition of relatively formal understanding would be "that understanding exhibited by Ss who, by common consent, should qualify as relatively formal thinkers," e.g., "fast" adolescent math students. This definition, a plausible guideline, confirmed that even high level compensation responses (there were in fact 6) were not formal solutions to the first spheres task since C responses (of any level) were far less frequent among abler Ss (and older Ss).

This study has illustrated some of the simpler levels of understanding controlling variables. Other levels, lower and higher, are the subject of a companion survey. There does not appear to be a way to uniquely define Piaget's formal stage and it is probably not useful to think there is a single "formal" concept of controlling variables and that it belongs wholly to a formal stage as opposed to a concrete one. Like all concepts, it can be more or less general. It has its developmental roots and these should not be neglected in the primary grades. Nor should development be taken for granted in the secondary grades, as ample evidence attests.

References

Case, R. (1975), "Structures and Strictures: Some Functional Limitations on the Course of Cognitive Growth," Cognitive Psychology (in press).

Dewey, J. (1971), How We Think, Henry Regnery Co., Chicago, 1971 (1st Ed. published in 1910).

Elkind, D. (1972), "Piaget and Science Education," Washington Monthly, 10, 3, 1972.

Hall, E. (1970), "A Conversation with Jean Piaget and Barbel Inhelder," Psychology Today, 3, 25, 1970

Inhelder, B. and Piaget, J. (1958), The Growth of Logical Thinking, Basic Books, 1958.

Karplus, R. and Karplus, E. (1972), "Intellectual Development Beyond Elementary School III--Ratio: A Longitudinal Study," School Science and Mathematics, 72, 9, 1972.

Kohlberg, L. and Gilligan, C. (1971), "The Adolescent as a Philosopher: The Discovery of the Self in a Postconventional World," Daedalus, 100, 1971.

Lawson, A. and Wollman, W. (1975), "Encouraging the Transition from Concrete to Formal Thinking: An Experiment," (in preparation).

Lunzer, E. (1973), "Formal Reasoning: A Re-Appraisal," 1973 Proceedings of the Jean Piaget Society, Temple University (in press).

Peel, E. (1971), The Nature of Adolescent Judgment, Staples, London, 1971 (Wiley-Interscience, New York).

Piaget J. (1932), The Moral Judgment of the Child, Free Press, 1932.

Renner, J. and Lawson, A (1973), "Promoting Intellectual Development Through Science Teaching," Physics Teacher 11, 1973.

Wollman, W. (1975), "Controlling Variables at a Concrete Stage" (in preparation).